A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* | Yumiko Maria Bejarano Azogue |
| *Student Number* | 2024144 |
| *Module Title* | Statistical Techniques for Data Analysis |
| *Assessment Title* | CA 1 – Capstone Project Proposal |
| *Assessment Due Date* | 21st April 2024 |
| *Date of Submission* | 21st April 2024 |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Contents

[Task 1 - Data Analysis: 3](#_Toc164302635)

[**Irish Gender Pay Gap** 3](#_Toc164302636)

[Conclusion 3](#_Toc164302637)

[Task 2 - Probability (Discrete): 4](#_Toc164302638)

[Question 4](#_Toc164302639)

[Answers 4](#_Toc164302640)

[Task 3 Probability (Continuous): 5](#_Toc164302641)

[Question 5](#_Toc164302642)

[Answers 5](#_Toc164302643)

[References 5](#_Toc164302644)

# Task 1 - Data Analysis:

## **Irish Gender Pay Gap**

### Characterization of Data

* Data Description: Begin by describing the dataset, including its size, structure, and variables. This involves examining the types of variables (numeric or categorical), their distributions, and any missing or erroneous values.
* Summary Statistics: Calculate descriptive statistics for numerical variables (e.g., mean, median, standard deviation) and frequency tables for categorical variables. This provides an initial understanding of the data's central tendencies and variability.

### Pre-processing

* Handling Missing Values: Address any missing data by imputation or removal, depending on the extent and nature of the missingness.
* Encoding Categorical Variables: Convert categorical variables into a numerical format suitable for analysis, such as one-hot encoding or label encoding.
* Normalization/Standardization: Scale numerical features to a similar range to prevent variables with larger magnitudes from dominating the analysis.

## Conclusion

In conclusion, this study sheds …...

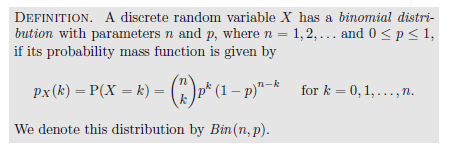
# Task 2 - Probability (Discrete):

### Question 1

What is the probability of rolling exactly two 6s in five rolls of a fair die?

### Answer

Use the binomial probability



​Where:

 = number of occurrences of a specific outcome in n trials

= probability of success in a single trial

 = number of trials

= number of combinations

Substituting the given values:

Binomial Coefficient

So, the probability of rolling exactly two 6s in five rolls of a fair die is approximately 0.160751

### Question 2

The number of industrial injuries on average per working week in a factory is 0.75. Assuming that the distribution of injuries follows a Poisson distribution, find the probability that in a particular week there will be no more than two accidents

### Answer

Use the binomial probability

….

# Task 3 Probability (Continuous):

## Question

The time a person spends at Dublin Zoo is Normally distributed with a mean of 90 minutes and a

standard deviation of 10 minutes.

Using this distribution, answer the following:

* If a visitor is selected at random, find the probability that they will spend at most 85 minutes visiting the zoo.
* If a visitor is selected at random, find the probability that they will spend at least 100 minutes visiting the zoo.
* Given that you know that a particular visitor has spent longer than average visiting the Zoo, what is the probability that they have spent more than 100 minutes there?

## Answers

Df……

# References

F.M., D., C., K., H.P. , L. and L.E. , M. (2005). *A modern introduction to probability and statistics : understanding why and how*. New York: Springer, Cop.

Keane, J. (n.d.). *First Derivatives (FD Technologies) Irish Gender Pay Gap Report*. [online] paygap.ie. Available at: https://gpg.ie/company/fdtechnologies [Accessed 18 Apr. 2024]. Dataset.

Keane, J. (n.d.). *Ireland’s Gender Pay Gap Reporting – Jennifer Keane*. [online] jenniferkeane.ie. Available at: https://jenniferkeane.ie/irelands-gender-pay-gap-reporting/ [Accessed 18 Apr. 2024].

Keane, J. (n.d.). *Understanding The Data - Irish Gender Pay Gap Portal*. [online] paygap.ie. Available at: https://paygap.ie/understandingTheData [Accessed 18 Apr. 2024].

OpenAI (2024). *ChatGPT*. [online] chat.openai.com. Available at: https://chat.openai.com/ Get instant answers, find inspiration, learn something new.

www.gov.ie. (2022). *What is the Gender Pay Gap Information Act 2021?* [online] Available at: https://www.gov.ie/en/publication/29606-what-is-the-gender-pay-gap-information-act-2021/.